

Laid-Open No.: 1997-0002853

Title: Voice recognition method using variable information rate model and information rate learning method

Abstract

Present research relates to a voice recognition method using Hidden Markov Model (HMM) and, more particularly, to a voice recognition method using a variable information rate model in which significance of each state is differentiated and an information rate learning method. The characteristics of the voice recognition method of the present research are that the method uses HMM which obtains a probability value  $Pr$  of a word model  $\lambda$  modeled from a discrete observation row  $O=\{O_1, O_2, \dots, O_T\}$  in the HMM method, but the states that form the word model  $\lambda$  contribute to the probability value  $Pr$  unevenly. The voice recognition method using HMM can recognize voice efficiently by giving a probability value a different weight according to the significance of each state.

(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(51) Int. Cl. <sup>6</sup> G10L 15/00		(45) 공고일자 2003년02월 11일
		(11) 등록번호 10-0355394
		(24) 등록일자 2002년09월 23일
(21) 출원번호 10-1995-0018112	(65) 공개번호 특 1997-0002853	
(22) 출원일자 1995년06월 29일	(43) 공개일자 1997년01월 28일	
(73) 특허권자 삼성전자 주식회사		
(72) 발명자 경기 수원시 팔달구 매탄3동 416 김남수 서울특별시서초구양재동 154-2우성아파트 102-1101 김경선 인천광역시남동구구월1동 1232-6 권석훈, 이영필, 오규환		
(74) 대리인		

심사관 : 남인호

(54) 가변정보율모델을이용한음성인식방법및정보율학습방법

요약

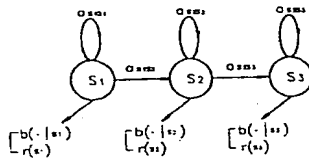
히든마코프모델(HMM: Hidden Markov Model)을 이용한 음성인식방법에 관한 것으로서 더욱 상세하게는 각 스테이트의 중요도를 달리한 가변정보율모델을 이용한 음성인식방법 및 정보율을 학습시키는 방법에 관한 것이다.

본 발명에 따른 음성인식방법은 이산화된 관측열  $O = \{O_1, O_2, \dots, O_T\}$ 로부터 HMM방법에 의해 모델링된 단어모델  $\lambda$ 의 확률값  $Pr$ 을 구하는 HMM을 이용한 음성인식방법에 있어서, 상기 단어모델  $\lambda$ 을 구성하는 각각의 스테이트가 확률값  $Pr$ 에 기여하는 정도를 균일하게 하지 않는 것을 특징으로 한다.

본 발명에 따른 HMM을 이용한 음성인식방법에 있어서는 각 스테이트마다의 중요도에 따라 확률값이 가중치를 갖게 함으로써 효율적인 음성인식을 달성하는 효과를 갖는다.

대표도

제 2 도



명세서

도면의 간단한 설명

제1도는 일반적인 음성인식장치의 구성을 보이는 블록도이다.

제2도는 가변정보율을 수용한 3스테이트로 구성되는 HMM모델의 일예를 보이는 도면이다.

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 히든마코프모델(HMM: Hidden Markov Model)을 이용한 음성인식방법에 관한 것으로서 더욱 상세하게는 각 스테이트의 중요도를 달리한 가변정보율모델을 이용한 음성인식방법 및 정보율을 학습시키는 방법에 관한 것이다.

현재 음성인식은 가장 자연스러운 맨-머신 인터페이스(man-machine interface)수단으로 알려져 있다. 음성인식방법으로는 크게 패턴정합을 이용한 동적프로그래밍 기법을 이용한 DTW(Dynamic Time Warping) 기법, 음성을 확률적으로 모델링하는 HMM방법이 있다. DTW방법은 인식대상어휘가 적을 때는 높은 인식률을

얻을 수 있지만 인식대상어휘가 많아 질수를 인식시간이 많이 소요되고 연속음의 경우에는 적용하기 어렵다는 단점이 있다.

이에 비해 HMM방법은 기준패턴을 만드는 데는 많은 시간을 필요로 하지만 인식시간이 매우 짧고, 인식을 이 DTW와 비슷하며, 화자적응방법을 쉽게 적용할 수 있다는 장점이 있어서, 최근에 많이 사용되고 있다.

HMM은 음성발생을 몇개의 상태(state)로 구성되는 Markov Model로 나타내고 각 state에서 출력발생을 확률분포로 특성짓는다.

HMM을 이용한 음성인식방법은 주어진 음성입력에 대하여 각 단어모델의 확률값을 구하고, 이중에서 가장 높은 확률값을 주는 단어를 선택한다.

이와 같이 성능이 우수한 HMM이지만 실제 적용에 있어서는 몇가지의 문제점을 내포하고 있다. 특히, 음성분석에 있어서 고정된 정보만을 이용하기 때문에 상대적으로 길이가 긴 모음구간은 중요시되지만 길이가 짧은 자음구간은 무시되는 경향이 있다.

실제에 있어서는 변화가 심한 자음구간에 더 중요한 인식의 단서가 있는 경우가 많다. 따라서, 각 부분의 인식정보의 중요도에 따라 음성분석률을 달리하는 것이 필요하다.

본 발명은 상기의 요구에 부응하기 위하여 창출된 것으로서 HMM의 각 state에서 가변정보율을 수용하도록 하는 음성인식방법을 제공하는 것을 그 목적으로 한다.

본 발명의 다른 목적은 상기의 정보율을 학습시키는 방법을 제공하는 것에 있다.

상기의 목적을 달성하는 본 발명에 따른 음성인식방법은

이산화된 관측열  $O = \{O_1, O_2, \dots, O_T\}$ 로부터 HMM방법에 의해 모델링된 단어모델  $\lambda$ 의 확률값  $Pr$ 을 구하는 HMM을 이용한 음성인식방법에 있어서,

상기 단어모델  $\lambda$ 를 구성하는 각각의 스테이트가 확률값  $Pr$ 에 기여하는 정도를 균일하게 하지 않는 것을 특징으로 한다.

상기의 다른 목적을 달성하는 본 발명에 따른 학습방법은

각 스테이트가 실제단어의 인식에 기여하는 확률값  $Pr$ 에 기여하는 정도(정보율)가 다르게 설정된 HMM모델의 정보율을 학습시키는 방법에 있어서,

학습에 사용된 샘플과 샘플과 일치되는 실제단어와의 상호정보를 최대가 되도록 학습시키는 것을 특징으로 한다. 이하 첨부된 도면을 참조하여 본 발명을 상세히 설명한다.

제1도는 일반적인 음성인식 시스템을 보이는 블록도이다. 제1도에 도시된 장치는 입력된 음성데이터를 부호화하고 특징파라미터를 출력하는 음성처리부(10), 음성처리부(10)에서 제공되는 특징파라미터와 정해진 인식모델을 이용하여 각 모델파라미터를 추정하는 학습부(12), 학습부(12)에서 추정된 각 모델파라미터를 저장하는 파라메타 저장부(14), 파라메타 저장부(14)에 저장된 모델파라미터를 이용하여 입력된 음성에 대한 인식을 수행하는 인식부(16)를 구비한다.

음성처리부(10)에서 제공되는 관측열을  $O = (O_1, O_2, \dots, O_T)$ 라 할 때, 단어모델  $\lambda$ 가 발생할 확률값  $Pr(O | \lambda)$ 는 다음과 같다.

$$Pr(O | \lambda) = \sum_x Pr(O, x | \lambda) \cdot Pr(x) \quad \dots \dots (1)$$

여기서,  $x$ 는 하나의 스테이트시퀀스를 나타내고,  $Pr(x)$ 는  $x$ 의 확률값을 나타낸다.

일반적으로 음성인식에서는 가장 높은 확률값을 주는 스테이트 시퀀스 하나만을 고려하므로

$$Pr(O | \lambda) \cong Pr(O, x_\lambda | \lambda) : Pr(x_\lambda) = Pr(x_\lambda) \prod_{i=1}^T Pr(O_i | x_{\lambda,i}) \quad \dots \dots (2)$$

로 주어질 수 있다.

여기서,  $x_\lambda$ 는 가장 높은 확률값을 주는 스테이트 시퀀스이고,  $x_{\lambda,i}$ 는  $x_\lambda$ 의 시간  $t$ 에서의 스테이트를 나타낸다.

제2도는 본 발명에 따른 가변정보율을 수용한 3개의 스테이트를 갖는 left-to-right HMM모델의 일 예를 보이는 도면이다. 제2도에 있어서  $\alpha_{s_i s_j}$ 는 스테이트  $S_i$ 로부터 스테이트  $S_j$ 로 전이하는 전이확률을 나타내고,  $b(\cdot | S_i)$ 는 스테이트  $S_i$ 에서의 출력확률분포를 나타내며, 그리고  $r(S_i)$ 는 스테이트  $S_i$ 에서의 정보율을 나타낸다.

본 발명의 가변정보율을 이용한 음성인식방법은 각 스테이트마다 가중치를 부여하기 위하여  $Pr(O_i | x_{\lambda,i})$ 를 정보율 파라메타  $r(x_{\lambda,i})$ 를 고려한 값으로 수정하여 사용한다.

$$Pr(O_i | x_{\lambda,i}) = [b(O_i | x_{\lambda,i})]^{r(x_{\lambda,i})} \quad \dots \dots (3)$$

여기서,  $b(O_i | x_{\lambda,i})$  스테이트  $x_{\lambda,i}$ 에서 출력  $O_i$ 를 발생시킬 확률을 나타내고,  $r(x_{\lambda,i})$ 는  $x_{\lambda,i}$ 에서의 정보율을 나타낸다.

정보를 파라메타  $r(\cdot)$ 는 각 스테이트의 중요도를 반영하며, 다음과 같은 조건을 만족한다.

$$\sum_{i=1}^M r(S_i) = D \cdot M \quad (4)$$

여기서,  $S_i$ 는 각 스테이트를 나타내고,  $M$ 은 전체 스테이트의 수, 그리고  $D$ 는 규정의 상수를 의미한다. 본 발명의 실제 실험에 있어서는  $D$ 가 1 - 10의 사이에 있을 때 인식성능의 향상이 있었다.

식(3)에 보여지는 정보율을 학습하는 방법은 다음과 같다.

학습부(12)에서는 기존의 HMM을 이용한 인식방법에서의 파라메타 학습과는 별도로 정보를 파라메타  $r(\cdot)$ 의 학습이 이루어진다.

먼저, 상호정보량  $I$ 는 주어진 음성과 설정된 모델 사이의 관계도를 나타내는 척도로서

$$I = \sum_{O_w} K(O_w, \lambda_w) \cong \sum_{O_w} \log \left[ \frac{\Pr(O_w | \lambda_w)}{\sum_{k=1}^N \Pr(\lambda_k) \cdot \Pr(O_w | \lambda_k)} \right] \quad (5)$$

와 같이 정의된다.

여기서,  $O_w$ 는 학습데이터 중의 하나의 샘플을 의미하며,  $I(O_w, \lambda_w)$ 는 상호정보를 표시하며,  $\lambda_k$ 는 모든 단어모델 중에서  $O_w$ 에 대하여  $k$ 번째로 높은 확률값을 부여하는 단어모델을 뜻하며,  $\Pr(\lambda)$ 는 단어모델  $\lambda$ 의 선확률을 표시한다.

$I$ 가 크면 음성과 이를 특징지우는 모델사이의 관련도가 큰 것을 의미한다. 따라서  $r(\cdot)$ 의 학습은  $I$ 를 극대화하는 방향으로 이루어지는데 이를 maximum mutual information(MMI)방법이라 한다.

구체적인 MMI 학습방법은 다음과 같다. 우선 각 스테이트  $S$ 에서의 정보율  $r(S)$ 에 대한  $I$ 의 미분을 구한

$$\frac{\partial I}{\partial r(S)} = \sum_{O_w} \left\{ \frac{\partial}{\partial r(S)} [\log \Pr(O_w | \lambda_w)] - \sum_{k=1}^N \left[ \frac{\Pr(\lambda_k)}{\Pr(O_w)} \cdot \frac{\partial \Pr(O_w | \lambda_k)}{\partial r(S)} \right] \right\}$$

----- (6)

$$\text{여기서, } \Pr(O_w) = \sum_{k=1}^N \Pr(\lambda_k) \cdot \Pr(O_w | \lambda_k) \quad (7)$$

이다.

식(6)의 미분값을 사용하여  $r(\cdot)$ 를 학습하는 데는 다음의 두가지 방법이 사용된다. 첫번째는 steepest ascent(점진적 증가)방법을 사용하는 것으로서,  $r(S)$ 가

$$r(S) |_{t+1} = r(S) |_t + \Delta \cdot \frac{\partial I}{\partial r(S)} |_t \quad (8)$$

로 표현되는 바와 같이  $r(S)$ 를  $\Delta \cdot \frac{\partial I}{\partial r(S)}$  만큼 증가시켜 가면서 최적의 확률값을 제공하는  $r(S)$ 를 찾도록 학습시킨다.

여기서,  $r(S) |_t$ 는  $t$ 번째 반복과정에서의  $r(S)$ 를 나타내며,  $\Delta$ 는 각 반복과정에서의 스텝사이즈를 나타낸다.

두번째의 학습방법은 steepest ascent 학습방법보다 수렴속도를 향상시킬 수 있는 것으로서 extended Baum(확장된 Baum) 알고리즘을 사용하는 것으로서,  $r(S)$ 가

$$r(S) |_{t+1} = \frac{r(S) |_t \cdot \frac{\partial I}{\partial r(S)} |_{t+\epsilon}}{\sum_{i=1}^M r(S_i) |_t \cdot \frac{\partial I}{\partial r(S_i)} |_{t+\epsilon}} \cdot M \quad (10)$$

로 표현되는 바와 같이 재추정하는 방법이다.

여기서,  $\epsilon$ 은 0보다 큰 수로서 다음과 같이 구해진다.

$$\epsilon = \max_{1 \leq i \leq M} \left\{ -\frac{\partial I}{\partial r(S_i)} |_{t,0} \right\} + \delta \quad (11)$$

여기서,  $\delta$ 는 0보다 큰 상수로 실험적으로 선정된다.

상술한 바와 같이 본 발명에 따른 HMM을 이용한 음성인식방법에 있어서 각 스테이트마다의 중요도에 따라 확률값이 가중치를 갖게 함으로써 효율적인 음성인식을 달성하는 효과를 갖는다.

(57) 청구의 범위

청구항 1

이산화된 관측열  $O = \{O_1, O_2, \dots, O_T\}$ 로부터 HMM방법에 의해 모델링된 단어모델  $\lambda$ 의 확률값  $Pr$ 을 구하는 HMM을 이용한 음성인식방법에 있어서,  
상기 단어모델  $\lambda$ 을 구성하는 각각의 스테이트가 확률값  $Pr$ 에 기여하는 정도를 균일하게 하지 않는 것을 특징으로 하는 음성인식방법.

청구항 2

제1항에 있어서, 각각의 스테이트에서의 출력확률분포는 하기의 식에 의해 표현됨을 특징으로 하는 음성인식방법.

$$Pr(O_t | x_{\lambda,t}) = [b(O_t | x_{\lambda,t})]^{r(x_{\lambda,t})}$$

여기서,  $b(O_t | x_{\lambda,t})$ 는 스테이트  $x_{\lambda,t}$ 에서 출력  $O_t$ 를 발생시킬 확률을 나타내고,  $r(x_{\lambda,t})$ 에서의 정보율을 나타냄.

청구항 3

제2항에 있어서, 상기 정보율  $r(x_{\lambda,t})$ 는 하기의 조건을 만족함을 특징으로 하는 음성인식방법.

$$\sum_{i=1}^M r(S_i) = D \cdot M$$

여기서,  $S_i$ 는 각 스테이트를 나타내고,  $M$ 은 전체 스테이트의 수, 그리고  $D$ 는 규정의 상수를 나타냄.

청구항 4

제3항에 있어서, 상기의  $D$ 는 1 내지 10 사이에 존재하는 실수임을 특징으로 하는 음성인식방법.

청구항 5

제3항에 있어서, 상기의  $D$ 는 학습에 사용된 샘플  $O$ 과 샘플과 일치되는 실제단어  $W$ 와의 상호정보를 최대가 되도록 학습시키는 과정에 의해 학습된 것임을 특징으로 하는 음성인식방법.

청구항 6

제5항에 있어서, 상기의 정보율은 현재의 정보율과 상기 미분값에 소정의 스텝사이즈를 곱한 값을 더한 값을 반복적으로 적용시키는 반복적 증가방법에 의해 구해진 것임을 특징으로 하는 음성인식방법.

청구항 7

제5항에 있어서, 상기 정보율은 재추정알고리즘에 의해 구해진 것임을 특징으로 하는 음성인식방법.

청구항 8

각 스테이트가 실제단어의 인식에 기여하는 확률값  $Pr$ 에 기여하는 정도(정보율)가 다르게 설정된 HMM모델의 정보율을 학습시키는 방법에 있어서,  
학습에 사용된 샘플과 샘플과 일치되는 실제단어와의 상호정보를 최대가 되도록 학습시키는 것을 특징으로 하는 정보율 학습방법.

청구항 9

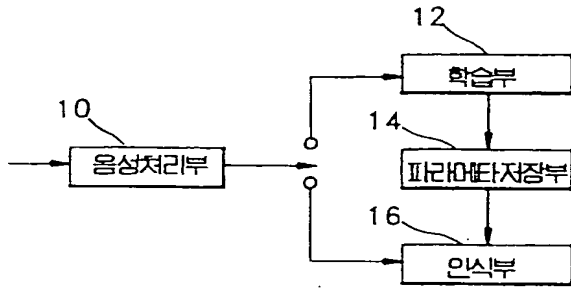
제8항에 있어서, 상기의 정보율은 현재의 정보율과 상기 미분값에 소정의 스텝사이즈를 곱한 값을 더한 값을 반복적으로 적용시키는 반복적 증가방법에 의해 구해진 것임을 특징으로 하는 음성인식방법.

청구항 10

제8항에 있어서, 상기 정보율은 재추정알고리즘에 의해 구해진 것임을 특징으로 하는 음성인식방법.

도면

도면1



도면2

